



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/00, A61K 39/215 C12N 5/00, C07K 13/00	A1	(11) International Publication Number: WO 86/ 05806 (43) International Publication Date: 9 October 1986 (09.10.86)
(21) International Application Number: PCT/GB86/00181	Needingworth, Huntingdon, Cambridgeshire PE17 3SA (GB); TOMLEY, Fiona, Margaret [GB/GB]; 56 Bateman Street, Cambridge CB2 1LR (GB).	
(22) International Filing Date: 27 March 1986 (27.03.86)	(74) Agent: PERCY, Richard, Keith; Patent Department, 101 Newington Causeway, London SE1 6BU (GB).	
(31) Priority Application Numbers: 8508265 8515678	(81) Designated States: DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, NL (European patent), US.	
(32) Priority Dates: 29 March 1985 (29.03.85) 20 June 1985 (20.06.85)	Published. With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(33) Priority Country: GB		
(71) Applicant (for all designated States except US): NATIONAL RESEARCH DEVELOPMENT CORPORATION [GB/GB]; 101 Newington Causeway, London SE1 6BU (GB)		
(72) Inventors; and (75) Inventors/Applicants (for US only): BINNS, Matthew, McKinley [GB/GB]; Appleby Cottage, Earsh Road, Colne, Huntingdon, Cambridgeshire PE17 5NJ (GB). BOURSNEIL, Michael, Edward, Griffith [GB/GB]; 11 Kent Close, St. Ives, Huntingdon, Cambridgeshire PE17 6DH (GB). BROWN, Thomas, David, Kay [GB/GB]; The Ferns, 3 High Street,		
(54) Title: INFECTIOUS BRONCHITIS VIRUS SPIKE PROTEIN		
(57) Abstract		
<p>The problem of diagnosis and typing of infectious bronchitis virus in poultry has been solved and important progress made towards an IBV vaccine by this invention. DNA complementary to the region of genomic IBV RNA which codes for a spike protein polypeptide comprising the S1 polypeptide (containing antigenic determinants) or the S2 polypeptide (containing means for anchoring the spike protein to the viral membrane) has been made. It can be carried by a cloning vector, incorporated in a host and cloned. It can also be cloned in a poxvirus which is used to transfect mammalian cells. Such cells express an artificial spike protein polypeptide.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT Austria	GA Gabon	MR Mauritania
AU Australia	GB United Kingdom	MW Malawi
BB Barbados	HU Hungary	NL Netherlands
BE Belgium	IT Italy	NO Norway
BG Bulgaria	JP Japan	RO Romania
BR Brazil	KP Democratic People's Republic of Korea	SD Sudan
CF Central African Republic	KR Republic of Korea	SE Sweden
CG Congo	LI Liechtenstein	SN Senegal
CH Switzerland	LK Sri Lanka	SU Soviet Union
CM Cameroon	LU Luxembourg	TG Togo
DE Germany, Federal Republic of	MC Monaco	US United States of America
DK Denmark	MG Madagascar	
FI Finland	ML Mali	
FR France		

- 1 -

INFECTIOUS BRONCHITIS VIRUS SPIKE PROTEINBackground of the invention1. Field of the invention

This invention relates to the spike protein of infectious
bronchitis virus (IBV) and to a recombinant DNA method or
05 preparing it. IBV is a virus which causes respiratory disease in
the fowl, and is of particular importance in relation to poultry.

2. Description of the prior art

IBV is a virus of the type Coronaviridae. It has a single-
stranded RNA genome, approximately 20 kb in length, of positive
10 polarity, which specifies the production of three major structural
proteins: nucleocapsid protein, membrane glycoprotein, and spike
glycoprotein. The spike glycoprotein is so called because it is
present in the teardrop-shaped surface projections or spikes
protruding from the lipid membrane of the virus. The spike protein
15 is believed likely to be responsible for immunogenicity of the
virus, partly by analogy with the spike proteins of other corona-
viruses and partly by in vitro neutralisation experiments, see,
for example, D. Cavanagh et al., Avian Pathology 13, 573-583 (1984).
Although the term "spike protein" is used to refer to the glycopro-
20 teinaceous material of the spike, it has recently been characterised
by D. Cavanagh, Journal of General Virology 64, 1187-1191; 1787-1791;
and 2577-2583 (1983) as comprising two or three copies each of two
glycopolypeptides, S1 (90,000 daltons) and S2 (84,000 daltons).
The polypeptide components of the glycopolypeptides S1 and S2 have
25 been estimated after enzymatic removal of oligosaccharides to have
a combined molecular weight of approximately 125,000 daltons. It
appears that the spike protein is attached to the viral membrane
by the S2 polypeptide.

The genomic organisation of the IBV viral proteins is
30 summarised in, for example, T.D.K. Brown and M.E.G. Boursnell,
Virus Research 1, 15-24 (1984). Briefly, six polyadenylated IBV
viral mRNA species (A to F) have been detected in infected cells.

- 2 -

mRNA A is the smallest and mRNA F is of genome length. These mRNAs form a so-called 'nested' or 3' co-terminal set. The nested mRNAs A to E have sizes approximately 2, 2.4, 3.4, 4.1 and 7.8 kb, as estimated from formaldehyde-agarose gel electrophoresis. They are shown in the accompanying drawing. Evidence from translation studies in vitro suggests that mRNAs A, C and E are each translated to give a corresponding major polypeptide. Thus, mRNA A codes for the nucleocapsid polypeptide, mRNA C for the membrane polypeptide and mRNA E for the precursor of the spike protein. In connection with mRNA E D.F. Stern and B.M. Sefton, Journal of Virology 50, 22-29 (1984) found that this mRNA specified production of the spike protein precursor in an in vitro translation. The sizes of the translation products are consistent with the coding capacity being present at the 5' end of each mRNA, but not present in the next smallest mRNA. In other words, the coding portion is within the "unique" region, i.e. the region of 'non-overlap' between successive RNAs of the set. U.v. inactivation studies have demonstrated that the subgenomic mRNAs are not produced by processing of larger RNA species, but are synthesised independently.

DNA complementary to IBV RNA (hereinafter referred to as 'cDNA') has been obtained for the Beaudette strain of IBV, as two fragments, together encompassing the first 3.3 kb of RNA from the 3' end, extending nearly to the 5' end of mRNA C. The fragments were inserted in plasmids and cloned in E. coli. They are described as C5.136 and C5.322 in T.D.K. Brown and M.E.G. Boursnell, supra, C5.136 being that running from nucleotides 1000 to 3300 approximately. Sequence information on C5.136 from nucleotides 1630 to 2400 approximately and the cloning of cDNA for IBV Beaudette strain including mRNA B and the 5' region of mRNA A have been described by M.E.G. Boursnell and T.D.K. Brown, Gene 29, 87-92 (1984). Further C5.136 sequence from nucleotides 2200 to 3400 approximately has been published by M.E.G. Boursnell, T.D.K. Brown and M.M. Binns, Virus Research 1, 303-313 (1984).

- 3 -

In the paper 'Genetically Engineered Vaccine against Avian Infectious Bronchitis Virus with the Advantages of Current Live and Killed Vaccines', by D. Cavanagh and the present inventors (M.M. Binns, M.E.G. Boursnell and T.D.K. Brown) in 'Modern Approaches to Vaccines', Cold Spring Harbor Laboratory, New York 1984, pages 215-218, it was announced that an oligonucleotide primer had been made and was currently being used to extend the C5.136 DNA so as to encompass the spike protein precursor gene. The oligonucleotide primer was described as corresponding to a sequence of 13 nucleotides approximately 150 bases in from the 5' terminus of C5.136. The nature and exact location of the oligonucleotide in the C5.136 cDNA sequence in the region from nucleotides 2400 to 3300 (the 5' terminus) have not been disclosed by these workers in any way, in writing or orally.

Summary of the invention

The present invention arises out of the research projected in broad outline above in 'Modern Approaches to Vaccines'. cDNA has been prepared by the primer method outlined above and within this cDNA sequences coding for the spike protein precursor (S) as well as sequences coding specifically for the S1 and S2 polypeptides have been identified. Cloned S, S1 and S2 DNA are starting materials for preparation of artificial polypeptides useful in a vaccine against IBV. Additionally, such DNA can be labelled to provide probes diagnostically useful in identifying IBV infections or in typing an infecting virus.

The research described has been carried out on three strains of IBV namely the Beaudette, M41 and 6/82 strains of IBV, but it is expected that other IBV serotypes and strains will exhibit a high degree of homology with one or more of these in respect of the spike protein precursor-coding cDNA.

According to an important feature of the invention there is provided a DNA molecule which codes for an IBV spike protein polypeptide comprising (consisting of or including) the S1 or S2 polypeptide. Such DNA is conveniently referred to as "spike DNA" for brevity. It includes DNA coding for the spike protein precursor.

- 4 -

Preferably there is at least 80%, more preferably at least 90%, amino acid sequence homology between the sequence coded for and the amino acid sequence of the corresponding polypeptide of the IBV Beaudette, M41 or 6/82 strain.

- 05 The invention includes specifically a DNA molecule which codes substantially only for any of (1) the spike protein precursor, (2) the S1 signal plus the S1 polypeptide, (3) the S1 polypeptide and (4) the S1 plus the S2 polypeptides, each said coding being to an extent of at least 80%, preferably at
- 10 least 90%, amino acid sequence homology between the sequence coded for and the amino acid sequence of the corresponding protein of the IBV Beaudette M41, or 6/82 strain.

- According to a preferred aspect of the invention there is included spike DNA as defined above which also shows at least 75%,
- 15 preferably at least 80%, more preferably at least 90%, and most preferably at least 95%, nucleotide sequence homology with the corresponding nucleotide sequence of the IBV Beaudette, M41 or 6/82 strain.

- In referring to DNA defined as coding substantially only for the various polypeptides it will be appreciated that it is intended
- 20 not to exclude flanking DNA sequences, which may be, for example, cDNA to flanking sequences in the IBV RNA genome or may be foreign sequences derived from other genes. Also, it is not intended that the S1 DNA should necessarily code for amino acids extending right
- 25 up to each terminus. It is expected that it will be possible to obtain expression of S1 cDNA lacking say, up to 5 or even 10 of the amino acids (30 nucleotides) at either terminus.

- The invention also includes a vector containing the above-defined IBV spike DNA, including a cloning vector such as a
- 30 plasmid or phage or an expression vector, preferably a poxvirus vector, and a host containing the vector. Mammalian cells containing the IBV spike DNA, whether as naked DNA or contained in a vector, are also included. Further, the invention includes artificial spike protein polypeptide and its expression from mammalian cells.

- 5 -

Brief description of the drawings

Figure 1 is a map of genomic and messenger RNA or IBV Beaudette strain showing cDNA clones and a primer used in obtaining the spike DNA of this invention.

- 05 Figure 2 is a map of recombinant DNA which defines certain plasmids containing IBV spike cDNA of strains M41 and 6/82.

Description of the preferred embodiments

- Sequence formula (1) below shows the complete nucleotide sequence of a cDNA molecule of the invention obtained from IBV genomic RNA Beaudette strain. To appreciate more fully the
10 correspondence of this DNA with the genomic and mRNA it is useful first to refer to Figure 1 of the drawing which shows the nested set of mRNAs. Each mRNA has a "leader" sequence at its 5'-end, this being shown in the drawing as a small rectangle. The leader
15 sequence does not appear in the corresponding part of the genomic RNA, but only at the 5'-end of the whole genome. For convenience, we shall refer to the mRNA/genomic RNA common sequence as the "body" of the mRNA. The IBV spike protein precursor is located substantially wholly within that portion of mRNA E which
20 extends 5'-wards beyond the 5' terminus of the body of mRNA D on the genome, i.e. from approximately nucleotides 4000 to 7500 of the genome. Sequence formula (1) shows a cDNA extending from the 5' terminus of the body of the mRNA E on the genome at about nucleotide 39 to the two stop codons at nucleotides 3587 to 3592.
25 The start codon at nucleotides 101 to 103 begins an open reading frame of 3486 nucleotides coding for 1162 amino acids, indicating a non-glycosylated protein of molecular mass about 127,000 daltons.

- 6 -

GATTTGAGATTGAAAGCAACGCCAGTTGTTAATTTGAAAACTGAACAAAAGACAGACTTA
 10 20 30 40 50 60

GTCTTTAATTTAATTAAGTGTGGTAAGTTACTGGTAAGAGATGTTGGTAACACCTCTTTT
 70 80 90 100 110 120

L V T L L C A L C S A V L Y D S S S Y V
 ACTAGTGACTCTTTTGTGTGCACTATGTAGTGCTGTTTGTATGACAGTAGTTCTTACGT
 130 140 150 160 170 180

Y Y Y Q S A F R P P S G W H L Q G G A Y
 TTACTACTACCAAAGTGCCTTCAGACCACCTAGTGGTGGCATTACAAGGGGTGCTTA
 190 200 210 220 230 240

A V V N I S S E F N N A G S S S G C T V
 TCGGTAGTTAAACATTTCTAGCGAATTTAATAATGCAGGCTCTTCATCAGGGTGTACTGT
 250 260 270 280 290 300

G I I H G G R V V N A S S I A M T A P S
 TGGTATTATTCATGGTGTCTGCTGTTGTTAATGCTTCTTCTATAGCTATGACGGCACCGTC
 310 320 330 340 350 360

S G M A W S S S Q F C T A H C N F S D T
 ATCAGGTATGGCTTGGTCTAGCAGTCAGTTTGTACTGCACACTGTAATTTTTCAGATAC
 370 380 390 400 410 420

T V F V T H C Y K H G G C P L T G M L Q
 TACAGTGTGGTTACACATTGTTATAAACATGGTGGGTGTCCTTTAACTGGCATGCTTCA
 430 440 450 460 470 480

Q N L I R V S A M K N G Q L F Y N L T V
 ACAGAACTTATACGTCTTTCTGCTATGAAAAATGGCCAGCTTTTCTATAATTTAACAGT
 490 500 510 520 530 540

S V A K Y P T F R S F Q C V N N L T S V
 TAGTGTAGCTAAGTACCCTACTTTTAGATCATTTTCAGTGTGTTAATAATTTAACATCCGT
 550 560 570 580 590 600

Y L N G D L V Y T S N E T I D V T S A G
 ATATTTAAATGGTGATCTTGTGTACACCTCTAATGAGACCATAGATGTTACATCTGCAGG
 610 620 630 640 650 660

- 7 -

V Y F K A G G P I T Y K V M R E V K A L
TGTTTATTTTAAAGCTGGTGGACCTATAACTTATAAAGTTATGAGAGAAGTTAAAGCCCT
670 680 690 700 710 720

A Y F V N G T A Q D V I L C D G S P R G
GGCTTATTTTGTAAATGGTACTGCACAAGATGTTATTTTGTGTGATGGATCACCTAGAGG
730 740 750 760 770 780

L L A C Q Y N T G N F S D G F Y P F T N
CTGTAGCATGCCAGTATAATACTGGCAATTTTTCAGATGGCTTTTATCCTTTTACTAA
790 800 810 820 830 840

S S L V K Q K F I V Y R E N S V N T T C
TAGTAGTTTAGTTAAGCAGAAGTTTATTGTCTATCGTAAAACTAGTTAATACTACTTG
850 860 870 880 890 900

T L H N F I F H N E T G A N P N P S G V
TACGTTACACAATTTTCATTTTTCATAATGAGACTGGCGCCAACCTAATCCTAGTGGTGT
910 920 930 940 950 960

Q N I Q T Y Q T K T A Q S G Y Y N F N F
TCAGAATATTCAAACCTACCAAAACAAAACAGCTCAGAGTGGTTATTATAATTTTAATTT
970 980 990 1000 1010 1020

S F L S S F V Y K E S N F M Y G S Y H P
TTCCTTTTCGAGTAGTTTTGTTTATAAGGAGTCTAATTTTATGTATGGATCTTATCACCC
1030 1040 1050 1060 1070 1080

S C K F R L E T I N N G L W F N S L S V
AAGTTGTAATTTAGACTAGAACTATTAATAATGGCTTGTGGTTAATTCACTTTCAGT
1090 1100 1110 1120 1130 1140

S I A Y G P L Q G G C K Q S V F K G R A
TCAATTGCTTACGGTCCCTTCAAGGTGGTTGCAAGCAATCTGTCTTTAAAGGTAGAGC
1150 1160 1170 1180 1190 1200

T C C Y A Y S Y G G P S L C K G V Y S G
AACTTGTTGTTATGCTTATTCATATGGAGGTCCTTCGCTGTGTAAAGGTGTTTATTCAGG
1210 1220 1230 1240 1250 1260

E L D H N F E C G L L V Y V T K S G G S
TGAGTTAGATCATAATTTTGAATGTGGACTGTTAGTTTATGTTACTAAGAGCGGTGGCTC
1270 1280 1290 1300 1310 1320

- 8 -

R I Q T A T E P P V I T Q N N Y N N I T
 TCGTATACAAACAGCCAGTGAACCGCCAGTTATAACTCAAACAATTATAATAATATTAC
 1330 1340 1350 1360 1370 1380

L N T C V D Y N I Y G R T G Q G F I T N
 TTTAAACTTGTGTGTAGTTATAATATATGGCAGAACTGGCCAAGTTTATTACTAA
 1390 1400 1410 1420 1430 1440

V T D S A V S Y N Y L A D A G L A I L D
 TGTGACCGACTCAGCTGTTAGTTATAATTATCTAGCAGACGCGGTTTGGCTATTTTAGA
 1450 1460 1470 1480 1490 1500

T S G S I D I F V V Q G E Y G L N Y Y K
 TACATCTGGTTCCATAGACATCTTTGTTGTACAAGGTGAATATGGTCTTAATTATTATAA
 1510 1520 1530 1540 1550 1560

V N P C E D V N Q Q F V V S G G K L V G
 GGTTAACCCCTTGCAGAGATGTCAACCAGCAGTTTGTAGTTTCTGTGGTAAATTAGTAGA
 1570 1580 1590 1600 1610 1620

I L T S R N E T G S Q L L E N Q F Y I K
 TATTCTTACTTACAGTAATGAGACTGGTTCTCAGCTTCTTGAGAACCAAGTTTACATCAA
 1630 1640 1650 1660 1670 1680

I T N G T R R F R R S I T E N V A N C P
 AATCACTAATGGAACACGCTCGTTTTAGACGTTCTATTACTGAAAATGTTGCAAAATTGCC
 1690 1700 1710 1720 1730 1740

Y V S Y G K F C I K P D G S I A T I V P
 TTATGTTAGTTATGGTAAGTTTGTATAAAACCTGATGGCTCAATTGCCACAATAGTACC
 1750 1760 1770 1780 1790 1800

K Q L E Q F V A P L F N V T E N V L I P
 AAAACAATTGGAACAGTTTGTGGCACCTTTATTTAATGTTACTGAAAATGTGCTCATACC
 1810 1820 1830 1840 1850 1860

N S F N L T V T D E Y I Q T R M D K V Q
 TAACAGTTTCAACTTAAGTGTACAGATGAGTACATACAAACGCGTATGGATAAGGTCCA
 1870 1880 1890 1900 1910 1920

I N C L Q Y V C G S S L D C R K L F Q Q
 AATTAATTGCCTGCAGTATGTTTGTGGCAGTTCTCTGGATTGTAGAAAAGTTGTTCAACA
 1930 1940 1950 1960 1970 1980

- 9 -

Y G P V C D N I L S V V N S V G Q K E D
ATATGGGCCTGTTTTCGCGACAACATATTGCTCTGTAGTAAATAGTGTGGTCAAAAAGAAGA
1990 2000 2010 2020 2030 2040

M E L L N F Y S S T K P A G F N T P V L
TATGGAACCTTTTGAATTTCTATTCTCTACTAAACCGGCTGGTTTAAATACACCAGTTCTT
2050 2060 2070 2080 2090 2100

S N V S T G E F N I S L L L T N P S S R
TAGTAATGTTAGCACTGGTGAGTTTAATATTCTCTTCTGTAAACAACTCTAGTAGTCG
2110 2120 2130 2140 2150 2160

R K R S L I E D L L F T S V E S V G L P
TAGAAAGCGTTCTCTTATTGAAGACCTTCTATTACAAGCGTTGAATCTGTTGGAGTACC
2170 2180 2190 2200 2210 2220

T N D A Y K N C T A G P L G F F K D L A
AACAAATGACGCATATAAAAAATTGCACTGCAGGACCTTTAGGCTTTTAAAGGACCTTGC
2230 2240 2250 2260 2270 2280

C A R E Y N G L L V L P P I I T A E M Q
GTGTGCTCGTGAATATAATGGTTTGCTTGTGTTGCCTCCTATCATAACAGCAGAAATGCA
2290 2300 2310 2320 2330 2340

A L Y T S S L V A S M A F G G I T A A G
AGCTTTGTATACTAGTTCTCTAGTAGCTTCTATGGCTTTTGGTGATTACTGCAGCTGG
2350 2360 2370 2380 2390 2400

A I P F A T Q L Q A R I N H L G I T Q S
TGCTATACCTTTTGGCCACACAACCTGCAGGCTAGAATTAATCACTTGGGTATTACCCAGTC
2410 2420 2430 2440 2450 2460

L L L K N Q E K I A A S F N K A I G H M
ACTTTTGTGTAAGAATCAAGAAAAAATTGCTGCTTCCTTTAATAAGGCCATTGGTCATAT
2470 2480 2490 2500 2510 2520

Q E G F R S T S L A L Q Q I Q D V V S K
GCAGGAAGGTTTGAAGTAGACATCTCTAGCATTACAACAAATCAAGATGTTGTAGTAA
2530 2540 2550 2560 2570 2580

Q S A I L T E T M A S L N K N F G A I S
ACAGAGTGTATTCTTACTGAGACTATGGCATCACTTAATAAAAAATTTGGTGCTATTTC
2590 2600 2610 2620 2630 2640

- 10 -

S V I Q E I Y Q Q F D A I Q A N A Q V D
TTCTGTGATTCAAGAAATCTACCGCAATTTGACGCCATACAAGCAAATGCTCAAGTGGGA
2650 2660 2670 2680 2690 2700

R L I T G R L S S L S V L A S A K Q A E
TCGTCTTATAACTGGTAGATTGTCATCACTTTCTGTTTACCATCTGCTAAGCAGGCGGA
2710 2720 2730 2740 2750 2760

Y I R V S Q Q R E L A T Q K I N E C V K
GTATATTAGAGTGTCAACAGCGTGAGTTAGCTACTCAGAAAATTAATGAGTGTGTTAA
2770 2780 2790 2800 2810 2820

S Q S I R Y S F C G N G R H V L T I P Q
GTCACAGTCTATTAGGTACTCCTTTTGTGGTAATGGACGACATGTTCTAACCATACCGCA
2830 2840 2850 2860 2870 2880

N A P N G I V F I H F S Y T P D S F V N
AAATGCACCTAATGGTATAGTGTATACACTTTTCTTATACTCCAGATAGTTTGTGTTAA
2890 2900 2910 2920 2930 2940

V T A I V G F C V K P A N A S Q Y A I V
TGTTACTGCAATAGTGGGTTTTTGTGTAAAGCCAGCTAATGCTAGTCAGTATGCAATAGT
2950 2960 2970 2980 2990 3000

P A N G R G I F I Q V N G S Y Y I T A R
GCCCCTAATGGTAGGGGTATTTTTATACAAGTTAATGGTAGTTACTACATCACTGCAGC
3010 3020 3030 3040 3050 3060

D M Y M P R A I T A G D V V T L T S C Q
AGATATGTATATGCCAAGAGCTATTACTGCAGGAGATGTAGTTACGCTTACTTCTGTGCA
3070 3080 3090 3100 3110 3120

A N Y V S V N K T V I T T F V D N D D F
AGCAAATTATGTAAGTGTAATAAGACCGTCATTACTACATTCGTAGACAATGATGATTT
3130 3140 3150 3160 3170 3180

D F N D E L S K W W N D T K H E L P D F
TGATTTTAATGACGAATTGTCAAAATGGTGAATGATACTAAGCATGAGCTACCAGACTT
3190 3200 3210 3220 3230 3240

D K F N Y T V P I L D I D S E I D R I Q
TGACAAATTCAATTACACAGTACCTATACTTGACATTGATAGTGAATGATCGTATTCA
3250 3260 3270 3280 3290 3300

- 11 -

G V I Q G L N D S L I D L E K L S I L K
AGGCGTTATACAGGGTCTTAATGACTCTCTAATAGACCTTGAAAACTTTCAATACTCAA
3310 3320 3330 3340 3350 3360

T Y I K W P W Y V W L A I A F A T I I F
AACTTATATTAAGTGGCCTTGGTATCTGTGGTTAGCCATAGCTTTTGGCCACTATTATCTT
3370 3380 3390 3400 3410 3420

I L I L G W V F F M T G C C G C C C G C
CATCTTAATACTAGGATGGGTTTTCTTCATGACTGGTTGTTGTGGTTGTTGTTGTGGATG
3430 3440 3450 3460 3470 3480

F G I M P L M S K C G K K S S Y Y T T F
CTTTGGCATTATGCCTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTATTACAGCACTTT
3490 3500 3510 3520 3530 3540

D N D V V T E Q Y R P K K S V * *
TGATAACGATGTGGTAAGTGAACAATACAGACCTAAAAAGTCTGTTTGATGATCCAAAGT
3550 3560 3570 3580 3590 3600

CCCACGTCCTTCCTAATAGTATTAATTCTTCTTTGGTGTAACCTT
3610 3620 3630 3640

(1)

- 12 -

This molecular weight is close to that estimated for the polypeptide components S1 and S2. In vitro translation of mRNA E had indicated that the non-glycosylated spike precursor protein had a molecular weight of 110,000 daltons while estimates of the
10 combined molecular weight of S1 and S2 after the removal of oligosaccharides by endoglycosidase H were 115,000 and 125,000.

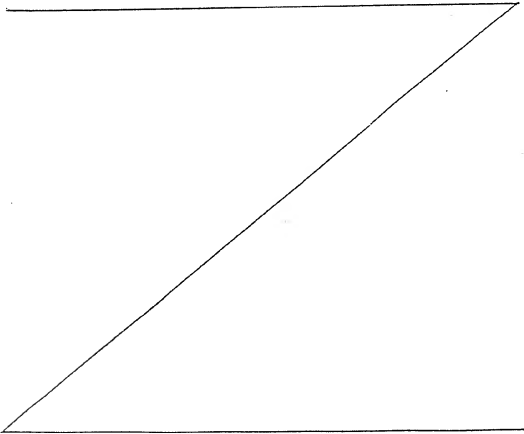
The cDNA contains sequences AACTGAACAAAA towards the 5' end and AACTGAACAATA towards the 3' end, underlined in formula (1). From their high homology with sequences, referred to in the
10 drawing as 'homology regions', which have previously been found at the 5' ends of the bodies of IBV mRNAs A, B and C and from mRNA length measurements it appears that these sequences represent approximately the position of the 5' ends of the bodies of mRNAs E and D. Surprisingly, the coding sequences for the spike protein
15 gene are not completely contained within the 'unique' region of mRNA E but extend for approximately 32 bases beyond the predicted 5' terminus of the body of mRNA D.

The spike protein precursor cDNA can be regarded as all that cDNA present in the open reading frame, including a signal region
20 from nucleotides 101 to 154 shown boxed, an S1 polypeptide-coding region from nucleotides 155 to 1696 and an S2 polypeptide-coding region from 1712 to 3586. The S1 and S2 polypeptide-coding regions are joined by a sequence from 1697 to 1711 coding for the amino acids RRFRR. This sequence of amino acids present in the precursor
25 polypeptide is believed to be cleaved during post-translation processing. The 5' end of the S2 sequence has been determined by amino acid sequencing and is shown arrowed at nucleotide 1712. Other features of the formula (1) sequence are referred to in the Examples hereinafter.

30 cDNA for spike protein polypeptides of the well known strain M41 and the strain 6/82 has been prepared using Beaudette strain RNA or cDNA as a hybridisation probe or to make a primer. Strain 6/82 was isolated in 1982 by Jane K.A. Cook, Vet. Record 112, 104-105 (1983) and is available without restriction from Houghton Poultry Research Station, Houghton, Huntingdon,

- 13 -

Cambridgeshire PE7 2DA, England, subject, of course, to compliance with legal regulations. Strain 6/82, which is isolate No. 2 of J.K.A. Cook, Avian Pathology 13, 733-741 (1984), exhibits cross-neutralisation reactions with Dutch serotypes.

- 05 Sequence formula (2) below compares the spike DNA sequences for Beaudette, M41, and 6/82, the 5'-end of which is the same for Beaudette as in sequence formula (1). There is a region of relatively high heterology between nucleotides 449 to 499, including particularly 458 to 463 for 6/82, which has six extra
- 10 nucleotides not present in M41 or Beaudette. The numbering system was therefore adjusted to align with 6/82, with the result that the Beaudette nucleotides after 458 are six numbers on from those in sequence formula (1). Overall, the three sequences show a high degree of homology. An analysis of M41 and Beaudette showed 70/3510
- 15 nucleotide changes resulting in 43/1139 amino acid changes. Strain 6/82 shows a lower degree of homology with M41 or Beaudette.
- 

- 14 -

	10	20	30	40	50
BEAU	GATTTGAGATTGAAAGCAACGCCAGTTGTTAATTTGAAAAGTGAACAAAA				
	60	70	80	90	100
BEAU	GACAGACTTAGTCTTTAATTTAATTAAGTGTGGTAAGTTACTGGTAAGAG				
	110	120	130	140	150
M41					ACTATGTAG
BEAU	ATGTTGGTAACACCTCTTTTACTAGTGACTCTTTTGTGTGCACATGTAG				
	160	170	180	190	200
M41	TGCTGCTTTTGTATGACAGTAGTTCTTACGTTTACTACTACCAAAGTGCCT				
BEAU	TGCTGTTTTGTATGACAGTAGTTCTTACGTTTACTACTACCAAAGTGCCT				
	210	220	230	240	250
M41	TTAGACCACCTAATGGTTGGCATTACACGGGGGTGCTTATGCGGTAGTT				
BEAU	TCAGACCACCTAGTGGTTGGCATTACAAAGGGGTGCTTATGCGGTAGTT				
	260	270	280	290	300
M41	AATATTTCTAGCGAATCTAATAATGCAGGCTCTTACCTGGGTGATTGT				
BEAU	AACATTTCTAGCGAATTTAATAATGCAGGCTCTTCATCAGGGTGTACTGT				
	310	320	330	340	350
M41	TGGTACTATTTCATGGTGGTGGTGTGTTAATGCTTCTTCTATAGCTATGA				
BEAU	TGGTATATTTCATGGTGGTGGTGTGTTAATGCTTCTTCTATAGCTATGA				
	360	370	380	390	400
6/82					GTACGGCT
M41	CGGCACCGTCATCAGGTATGGCTTGGTCTAGCAGTCAGTTTTGTACTGCA				
BEAU	CGGCACCGTCATCAGGTATGGCTTGGTCTAGCAGTCAGTTTTGTACTGCA				
	410	420	430	440	450
6/82	CACTGCAATTTTACTGATTTTGTAGTATTTGTTACACATTGCTATAAAAG				
M41	CACTGTAACCTTTTTCAGATACTACAGTGTGTTGTTACACATTGTTATAAATA				
BEAU	CACTGTAATTTTTCAGATACTACAGTGTGTTGTTACACATTGTTATAAACA				
	460	470	480	490	500
6/82	TGGTCATGGTTCATGTCCTTTTAACAGGCTCGATTCCACAGAATCATATTC				
M41	TGATGGG-----TGTCCTATAACTGGCATGCGTCAAAAGAAATTTTTTAC				
BEAU	TGGTGGG-----TGTCCTTTAACTGGCATGCTTCAACAGAACTCTTATAC				
	510	520	530	540	550
6/82	GTATTTCTGCTATGAAAAATAGCAGTTTGTGTTTATAAAGTTTACAGTTGCT				
M41	GTCTTTCTGCTATGAAAAATGGCCAGCTTTTCTATAATTTAAACAGTTAGT				
BEAU	GTGTTTCTGCTATGAAAAATGGCCAGCTTTTCTATAATTTAAACAGTTAGT				
	560	570	580	590	600
6/82	GTGACTAAATATCCTAGATTTAAGTCGCTTCAGTGTGTTAATAATATGAC				
M41	GTAGCTAAGTACCCTACTTTTAAATCATTTCAGTGTGTTAATAATTTAAC				
BEAU	GTAGCTAAGTACCCTACTTTTAGATCATTTCAGTGTGTTAATAATTTAAC				
	610	620	630	640	650
6/82	ATCTGTATACCTAAATGGCGATCTCGTTTTTACTTCTAAGCAGACTAAAG				
M41	ATCCGTATATTTAAATGGTGATCTTGTGTTTACACCTCTAATGACACCATAG				
BEAU	ATCCGTATATTTAAATGGTGATCTTGTGTTTACACCTCTAATGACACCATAG				

- 15 -

	660	670	680	690	700
6/82	ATGTTAGTGTGTCAG				
M41	ATGTTACATCTGCAGGTGTTTATTTTAAAGCTGGTGGACCTATAACTTAT				
BEAU	ATGTTACATCTGCAGGTGTTTATTTTAAAGCTGGTGGACCTATAACTTAT				
	710	720	730	740	750
M41	AAAGTTATGAGAGAAGTTAAAGCCCTGGCTTATTTTGTTAATGGTACTGC				
BEAU	AAAGTTATGAGAGAAGTTAAAGCCCTGGCTTATTTTGTTAATGGTACTGC				
	760	770	780	790	800
M41	ACAAGATGTTATTTTGTGTGATGGATCACCTAGAGGCTTGTTAGCATGCC				
BEAU	ACAAGATGTTATTTTGTGTGATGGATCACCTAGAGGCTTGTTAGCATGCC				
	810	820	830	840	850
6/82	GTATAATACTGGTAAATTTTTCAGATGGCTTTTATCCTTTTACTAATAGT				
M41	AGTATAATACTGGCAATTTTTCAGATGGCTTTTATCCTTTTATAAAGT				
BEAU	AGTATAATACTGGCAATTTTTCAGATGGCTTTTATCCTTTTACTAATAGT				
	860	870	880	890	900
6/82	AGTTTAGTTAAGGAAAAGTTTATTGTTTATCGTGAAAGTAGTGTTAACAC				
M41	AGTTTAGTTAAGCAGAAGTTTATTGTTTATCGTGAAAGTAGTGTTAATAC				
BEAU	AGTTTAGTTAAGCAGAAGTTTATTGTTTATCGTGAAAGTAGTGTTAATAC				
	910	920	930	940	950
6/82	TACTTTGGAGTTAACTAATTTTCACTTTTCTAATGTAAGTAATGCTACCC				
M41	TACTTTACGTTACACAATTTTCACTTTTCAATAGAGACTGGCGCCAAAC				
BEAU	TACTTTGACGTTACACAATTTTCACTTTTCAATAGAGACTGGCGCCAAAC				
	960	970	980	990	1000
6/82	CTAACACAGGGGGTGTCAGACCATTCATTAATCAAAACCATCAGCGCT				
M41	CTAATCCTAGTGGTGTTCAGAAATTTCAAACCTACCAAAACACAAACAGCT				
BEAU	CTAATCCTAGTGGTGTTCAGAAATTTCAAACCTACCAAAACAAAAACAGCT				
	1010	1020	1030	1040	1050
6/82	CAGAGTGGTTATTATAATCTTAATTTCTCCTTTCTGAGTAGTTTTATTTA				
M41	CAGAGTGGTTATTATAATTTTAAATTTTCTCCTTTCTGAGTAGTTTTATTTA				
BEAU	CAGAGTGGTTATTATAATTTTAAATTTTCTCCTTTCTGAGTAGTTTTATTTA				
	1060	1070	1080	1090	1100
6/82	TAAGCGCTCTGATTATGTAATGGGCTTTACCAACC				
M41	TAAGGAGTCTAATTTTATGTAATGGATCTTATACCCCAAGTTGTAATTTTA				
BEAU	TAAGGAGTCTAATTTTATGTAATGGATCTTATACCCCAAGTTGTAATTTTA				
	1110	1120	1130	1140	1150
M41	GACTAGAAAACATAAATAATGGCTTTGTTTAAATTCAGTTTCACTTTTCA				
BEAU	GACTAGAAAACATAAATAATGGCTTTGTTTAAATTCAGTTTCACTTTTCA				
	1160	1170	1180	1190	1200
M41	ATTGCTTACGGTCCCTCTTCAAGGTGGTTGCAAGCAATCTGTCTTTAAGG				
BEAU	ATTGCTTACGGTCCCTCTTCAAGGTGGTTGCAAGCAATCTGTCTTTAAGG				
	1210	1220	1230	1240	1250
M41	TAGAGCAACTTGTGTGTTATGCTTATTCATATGGAGGTCCCTTCGCTGTGTA				
BEAU	TAGAGCAACTTGTGTGTTATGCTTATTCATATGGAGGTCCCTTCGCTGTGTA				
	1260	1270	1280	1290	1300
M41	AAGGTGTTTATTCAGGTGAGTTAGATCTTAATTTGAATGTGGACTGTTA				
BEAU	AAGGTGTTTATTCAGGTGAGTTAGATCATAATTTGAATGTGGACTGTTA				

- 16 -

	1310	1320	1330	1340	1350
M41	GTTTATGTTACTAAGAGCGGTGGCTCTCGTATACAAACAGCCACTGAACC				
BEAU	GTTTATGTTACTAAGAGCGGTGGCTCTCGTATACAAACAGCCACTGAACC				
	1360	1370	1380	1390	1400
M41	GCCAGTTATAACTCGACACAATTATAATAATTACTTTAAATCTTGTG				
BEAU	GCCAGTTATAACTCAAAACAATTATAATAATTACTTTAAATCTTGTG				
	1410	1420	1430	1440	1450
M41	TTGATTATAATATATATGGCAGAAGCTGGCCAAGGTTTTATTACTAATGTA				
BEAU	TTGATTATAATATATATGGCAGAAGCTGGCCAAGGTTTTATTACTAATGTA				
	1460	1470	1480	1490	1500
M41	ACCGACTCAGCTGTTAGTTATAATTATCTAGCAGACGCCAGGTTTGGCTAT				
BEAU	ACCGACTCAGCTGTTAGTTATAATTATCTAGCAGACGCCAGGTTTGGCTAT				
	1510	1520	1530	1540	1550
6/82					GGTGAATATG
M41	TTTAGATACATCTGGTTCCATAGACATCTTTGTTGTACAAGGTGAATATG				
BEAU	TTTAGATACATCTGGTTCCATAGACATCTTTGTTGTACAAGGTGAATATG				
	1560	1570	1580	1590	1600
6/82	GTCTTAATTATTATAAAGTTAACCCTTGTGAGGATGTTAATCAGCAGTTT				
M41	GTCTTACTTATTATAAAGTTAACCCTTGCAGAGATGTCACCCAGCAGTTT				
BEAU	GTCTTAATTATTATAAAGTTAACCCTTGCAGAGATGTCACCCAGCAGTTT				
	1610	1620	1630	1640	1650
6/82	GTAGTTTCTGGTGGTAAATTAGTAGGTATTCTTACGTCACGTAATGAGAC				
M41	GTAGTTTCTGGTGGTAAATTAGTAGGTATTCTTACTTCACGTAATGAGAC				
BEAU	GTAGTTTCTGGTGGTAAATTAGTAGGTATTCTTACTTCACGTAATGAGAC				
	1660	1670	1680	1690	1700
6/82	TGGCTCGCAGCCTCTTGAAAACCAAGTTCTATATTAATCATTAAATGGAA				
M41	TGGTTCTCAGCTTCTTGAGAACCAAGTTTACATTAAAAATCACTAATGGAA				
BEAU	TGGTTCTCAGCTTCTTGAGAACCAAGTTTACATCAAAATCACTAATGGAA				
	1710	1720	1730	1740	1750
6/82	CTCGTCGTTCTAGACGCTCTATTACTGGGAATGTTACAAAATTGTCCTTAT				
M41	CACGTCGTTTTAGACGTTCTATTACTGAAAATGTTGCAAAATTGCCCTTAT				
BEAU	CACGTCGTTTTAGACGTTCTATTACTGAAAATGTTGCAAAATTGCCCTTAT				
	1760	1770	1780	1790	1800
6/82	GTTACTTATGGCAAGTTTTGTATAAAACCTGATGCTCAATTTCACACCC				
M41	GTTAGTTATGGTAAGTTTTGTATAAAACCTGATGGTTCAATTGGCCACAAT				
BEAU	GTTAGTTATGGTAAGTTTTGTATAAAACCTGATGGTTCAATTGGCCACAAT				
	1810	1820	1830	1840	1850
6/82	ACCAAAAGAATTGGAACATTTTGGGCACCTCTACTTAATGTAACCTG				
M41	AGTACCAAAACAATTGGAACAGTTTGTGGCACCTTTACTTAATGTTACTG				
BEAU	AGTACCAAAACAATTGGAACAGTTTGTGGCACCTTTATTTAATGTTACTG				
	1860	1870	1880	1890	1900
6/82	AAAATGTGCTCATACCTGACAGTTTAAATTTAACTGTTACAGATGAGTAC				
M41	AAAATGTGCTCATACCTAACAGTTTAAATTTAACTGTTACAGATGAGTAC				
BEAU	AAAATGTGCTCATACCTAACAGTTTAACTTAACTGTTACAGATGAGTAC				

- 17 -

	1910	1920	1930	1940	1950
6/82	ATACAAACGCGTATGGATAAGGTCCAAATTAATTGCGCTTCAGTATGTTTG				
M41	ATACAAACGCGTATGGATAAGGTCCAAATTAATTGCTGTCAGTATGTTTG				
BEAU	ATACAAACGCGTATGGATAAGGTCCAAATTAATTGCGCTGTCAGTATGTTTG				
	1960	1970	1980	1990	2000
6/82	CGGCAATTCTTTGGAGTGTAGAAAGTGTGTTTCAA				
M41	TGGCAATTCTCTGGATTGTAGAGATTTGTTTCAACAATATGGGCCTGTTT				
BEAU	TGGCAGTTCTCTGGATTGTAGAAAGTGTGTTTCAACAATATGGGCCTGTTT				
	2010	2020	2030	2040	2050
M41	GTGACAACATATTGCTCTGTAGTAAATAGTATTGGTCAAAAAGAAGATATG				
BEAU	GCGACAACATATTGCTCTGTAGTAAATAGTATTGGTCAAAAAGAAGATATG				
	2060	2070	2080	2090	2100
M41	GAACTTTTGAATTTCTATTCTTCTACTAAACCGGCTGGTTTTAATACACC				
BEAU	GAACTTTTGAATTTCTATTCTTCTACTAAACCGGCTGGTTTTAATACACC				
	2110	2120	2130	2140	2150
M41	ATTTCTTAGTAATGTTAGCACTGGTGAGTTTAATATTCTCTTCTGTTAA				
BEAU	AGTTCTTAGTAATGTTAGCACTGGTGAGTTTAATATTCTCTTCTGTTAA				
	2160	2170	2180	2190	2200
6/82	CAACTCCTAGTAGTCCTAGAAGGCGTTCCTTTATTGAAGACCTTCTATTT				C
M41	CAAATCCTAGTAGTCGTAGAAGGCGTTCCTTTATTGAAGACCTTCTATTT				
BEAU	CAAATCCTAGTAGTCGTAGAAGGCGTTCCTTTATTGAAGACCTTCTATTT				
	2210	2220	2230	2240	2250
6/82	ACAAGTGTGTAATCTGTTGGATTACCAACAGATGACGCATACAAAGAAGTG				
M41	ACAAGCGTGTGAATCTGTTGGATTACCAACAGATGACGCATACAAAAATTG				
BEAU	ACAAGCGTGTGAATCTGTTGGACTACCAACAAATGACGCATATAAAAAATTG				
	2260	2270	2280	2290	2300
6/82	CACCTGCAGGACCTTTAGGCTTTCTTAAGGACCTAGCGTGTGCTCGTGAAT				
M41	CACCTGCAGGACCTTTAGGCTTTCTTAAGGACCTTGCCTGTGCTCGTGAAT				
BEAU	CACCTGCAGGACCTTTAGGCTTTTTTAAGGACCTTGCCTGTGCTCGTGAAT				
	2310	2320	2330	2340	2350
6/82	ATAATGGTTTGCTTGTGTTGCGCTCCTATTATAACAGCAGAAATGCAAAACC				
M41	ATAATGGTTTGCTTGTGTTGCGCTCCTATTATAACAGCAGAAATGCAAAACC				
BEAU	ATAATGGTTTGCTTGTGTTGCGCTCCTATCATACAGCAGAAATGCAAGCT				
	2360	2370	2380	2390	2400
6/82	TTGTATACATAGTTCTCTAGTAGCTTCTATGGCTTTTGGTGGTATTACTTC				
M41	TTGTATACATAGTTCTCTAGTAGCTTCTATGGCTTTTGGTGGTATTACTGC				
BEAU	TTGTATACATAGTTCTCTAGTAGCTTCTATGGCTTTTGGTGGTATTACTGC				
	2410	2420	2430	2440	2450
6/82	AGCTGGTGCTATACCTTTTGGCCACACAACCTGCAGGCTAGAATTAATCATT				
M41	AGCTGGTGCTATACCTTTTGGCCACACAACCTGCAGGCTAGAATTAATCACT				
BEAU	AGCTGGTGCTATACCTTTTGGCCACACAACCTGCAGGCTAGAATTAATCACT				
	2460	2470	2480	2490	2500
6/82	TGGGTATACCCAGTCACCTTTGTTTGAAGAAATCAAGAAAAAATGCTGCT				
M41	TGGGTATTACCCAGTCACCTTTGTTTGAAGAAATCAAGAAAAAATGCTGCT				
BEAU	TGGGTATTACCCAGTCACCTTTGTTTGAAGAAATCAAGAAAAAATGCTGCT				

- 18 -

	2510	2520	2530	2540	2550
M41	TCCTTTAATAAGGCCATTGGTTCGTATGCAGGAAGGTTTTAGAAGTACATC				
BEAU	TCCTTTAATAAGGCCATTGGTCATATGCAGGAAGGTTTTAGAAGTACATC				
	2560	2570	2580	2590	2600
M41	TCTAGCATTACAAACAAATTCAAGATGTTGTTAATAAGCAGAGTGCTATTTC				
BEAU	TCTAGCATTACAAACAAATTCAAGATGTTGTTAGTAAACAGAGTGCTATTTC				
	2610	2620	2630	2640	2650
M41	TTACTGAGACTATGGCATCACTTAATAAAAAATTTTGGTGTCTATTCTCTCT				
BEAU	TTACTGAGACTATGGCATCACTTAATAAAAAATTTTGGTGTCTATTCTCTCT				
	2660	2670	2680	2690	2700
M41	GTGATTCAAGAAATCTACCAGCAACTTGACGCCATACAAGCAAATGCTCA				
BEAU	GTGATTCAAGAAATCTACCAGCAATTTGACGCCATACAAGCAAATGCTCA				
	2710	2720	2730	2740	2750
M41	AGTGGATCGTCTTATAACTGGTAGATTGTCATCACTTTCTGTTTTAGCAT				
BEAU	AGTGGATCGTCTTATAACTGGTAGATTGTCATCACTTTCTGTTTTAGCAT				
	2760	2770	2780	2790	2800
M41	CTGCTAAGCAGGCGGAGCATATTAGAGTGTCAACACAGCGTGAGTTAGCT				
BEAU	CTGCTAAGCAGGCGGAGTATATTAGAGTGTCAACACAGCGTGAGTTAGCT				
	2810	2820	2830	2840	2850
6/82	AAATTAATGAGTGTGTTAAATCTCAATCTATTAGGTATTCATT				
M41	ACTCAGAAAAATTAATGAGTGTGTTAAGTCACAGTCTATTAGGTACTCCTT				
BEAU	ACTCAGAAAAATTAATGAGTGTGTTAAGTCACAGTCTATTAGGTACTCCTT				
	2860	2870	2880	2890	2900
6/82	TTGTGGTAATGGAAGACATGTTCTAACCATACCACAAAAATGCTCCTAATG				
M41	TTGTGGTAATGGAAGACATGTTCTAACCATACCACAAAAATGCACTTAATG				
BEAU	TTGTGGTAATGGAAGACATGTTCTAACCATACCACAAAAATGCACTTAATG				
	2910	2920	2930	2940	2950
6/82	GCATAGTGTTTATACACTTTTACATACACGCCAGAGAGTTTTGTCAATGTG				
M41	GTATAGTGTTTATACACTTTTCTTATACTCCAGATAGTTTTGTAAATGTT				
BEAU	GTATAGTGTTTATACACTTTTCTTATACTCCAGATAGTTTTGTAAATGTT				
	2960	2970	2980	2990	3000
6/82	ACGGCAATAGTAGGGTTTTGTGTAACCCAGCTAATGCTAGCCAGTATGC				
M41	ACTGCAATAGTGGGTTTTGTGTAACCCAGCTAATGCTAGTCAGTATGC				
BEAU	ACTGCAATAGTGGGTTTTGTGTAACCCAGCTAATGCTAGTCAGTATGC				
	3010	3020	3030	3040	3050
6/82	AATAGTGCCTGCTAATGGCAGAGGTATTTTTATACAAGTTAATGGTAGTT				
M41	AATAGTACCCGCTAATGGTAGGGGTATTTTTATACAAGTTAATGGTAGTT				
BEAU	AATAGTACCCGCTAATGGTAGGGGTATTTTTATACAAGTTAATGGTAGTT				
	3060	3070	3080	3090	3100
6/82	ACTACATCACTGCAAGAGATATGTATATGCCAAGAGATATTACTGCAGGA				
M41	ACTACATCAACAGCAGAGATATGTATATGCCAAGAGCTATTACTGCAGGA				
BEAU	ACTACATCACTGCAAGAGATATGTATATGCCAAGAGCTATTACTGCAGGA				

- 19 -

6/82	3110	3120	3130	3140	3150
M41	GATATAGTTACGCTTACTTCTGTGTCAGCAAATATGTAAGTGTAATAA				
BEAU	GATATAGTTACGCTTACTTCTGTGTCAGCAAATATGTAAGTGTAATAA				
6/82	3160	3170	3180	3190	3200
M41	GACCGTCATTACTACATTTGTAGACAATGATGATTTTGATTTTGATGACG				
BEAU	GACCGTCATTACTACATTCGTAGACAATGATGATTTTGATTTTAATGACG				
6/82	3210	3220	3230	3240	3250
M41	AGTTGTCAAATGGTGGAAATGATACTAAGCATGAGCTACCAGACTTTTGAC				
BEAU	AATTGTCAAATGGTGGAAATGATACTAAGCATGAGCTACCAGACTTTTGAC				
6/82	3260	3270	3280	3290	3300
M41	GAATTC AATTATACAGTACCTATACTTGATATTGGTAGTGAAATTGATCG				
BEAU	AAATTC AATTATACAGTACCTATACTTGACATTGATAGTGAAATTGATCG				
6/82	3310	3320	3330	3340	3350
M41	TATTCAAGGTGTTATACAGGGCCTTAATGACTCTCTAATAGACCTTGAAA				
BEAU	TATTCAAGGCGTTATACAGGGTCTTAATGACTCTCTAATAGACCTTGAAA				
6/82	3360	3370	3380	3390	3400
M41	CCCTTTCAACTACTTAAGACTTATATTAATGGCCTTGATGTGTGGCTT				
BEAU	AACTTTCAACTACTCAAAACTTATATTAAGTGGCCTTGATGTGTGGTTA				
6/82	3410	3420	3430	3440	3450
M41	GCCATTGCATTCCTTACCATTATCTTTATTCTGGT				
BEAU	GCCATAGCTTTTGGCACTATTATCTTCATCTTAATACTAGGATGGGTTTT				
M41	3460	3470	3480	3490	3500
BEAU	CTTCATGACTGGATGTTGTGGTGTGTTGTGTTGGATGCTTTGGCATTATGC				
M41	3510	3520	3530	3540	3550
BEAU	CTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTATTACACGACTTTTGAT				
M41	3560	3570	3580	3590	3600
BEAU	AACGATGTGGTAACTTAACAATACAGACCTAAAAAGTCTGTTTAATGATT				
M41	3610	3620	3630	3640	3650
BEAU	CAAAGTCCCACGTCTCTCTAATAGTATTAATCTCTTTGGTGTAACCT				
M41	3660	3670	3680	3690	3700
BEAU	T				

(2).

- 20 -

The IBV RNA of many other strains is believed to be fairly similar to that of Beaudette, M41 or 6/82 and therefore DNA molecules of the present invention can be used as probes for hybridisation to RNA of other serotypes, thus enabling spike cDNA of other strains to be identified and prepared. For example, cDNA from other IBV strains of the Massachusetts serotype or the live vaccine strains H52 and H120 used in the UK, believed to be similar to M41 could be prepared from M41 or Beaudette cDNA. Any of the Dutch type strains in the serogroups known as D207, D212, D3128, and D3896, believed to be similar to strain 6/82 (Houghton Poultry Research Station, Huntingdon, England), could be prepared using 6/82 DNA and probably from M41 or Beaudette cDNA. Even if the overall degree of homology between any of these IBVs and the starting strain IBV (i.e. Beaudette, M41 or 6/82) is not high enough to allow hybridisation over a substantial length of sequence, it can confidently be expected that there will be some lengths of at least 13 nucleotides, and more desirably at least 18 nucleotides, which have very high homology, allowing series of such probes to be constructed from starting strain IBV spike cDNA. Some of these probes will hybridise to cDNA of the RNA of the other IBV. By probing a library of such cDNA, spike protein cDNA of the other IBV can be identified and obtained. Alternatively, the "random priming" method described above for preparation of M41 and 6/82 cDNA from Beaudette can be used to prepare cDNA from any strain.

The invention therefore also includes particularly DNA molecules coding for IBV spike protein polypeptide having a reasonable degree of homology of nucleotide sequence with IBV Beaudette, M41 or 6/82 strain to allow hybridisation to take place. The suggested minimum degree of nucleotide sequence homology is 75%, but at least 80% is preferred and a degree of homology of 85-100% would be useful in normally allowing hybridisation to take place under reasonably stringent conditions. (Obviously, if the DNA is to be used in typing viruses one would perform a probe hybridisation under far more stringent conditions).

- 23 -

host such as a bacterial host, especially E. coli or Bacillus species, or a yeast. For expression, mammalian cells can be transfected by the calcium phosphate precipitation method or transformed by a viral vector. Viral vectors include retroviruses and poxviruses such as fowlpox virus or vaccinia virus.

The IBV spike DNA can be introduced into the viral vector as follows. The spike DNA is inserted into a plasmid containing an appropriate poxvirus gene, such as the thymidine kinase gene of vaccinia virus, so that the insert interrupts the gene sequence.

10 A virus promotor is also introduced into the gene sequence in such a position that it will operate on the inserted spike DNA sequence. When the poxvirus and the plasmid recombinant DNA are co-transfected into a mammalian cell, homologous recombination takes place between the poxvirus gene, such as TK in vaccinia virus, and the same gene

15 present in the plasmid. Since the IBV spike DNA has thereby interrupted the poxvirus gene, viruses lacking the gene expression product, such as TK, are selected. Once such a recombinant virus vector has been thus constructed it can be used to introduce the IBV spike DNA directly into the desired host cells without the

20 need for any separate step of transfecting plasmid recombinant DNA into the cells.

With a view ultimately to obtaining expression of the recombinant virus in vivo, the preferred poxvirus is fowlpox virus. It may be that the inserted IBV DNA contains a sequence,

25 which, in the fowlpox vector, lead to premature termination of transcription. In this case, the spike DNA would have to be modified slightly by one or two nucleotides, thereby to allow transcription to proceed along the full length of the gene.

The vector can be introduced into any appropriate host by any

30 method known in recombinant DNA technology. Hosts include E. coli, Bacillus spp, mammalian cells, and yeasts. The method of introduction

- 24 -

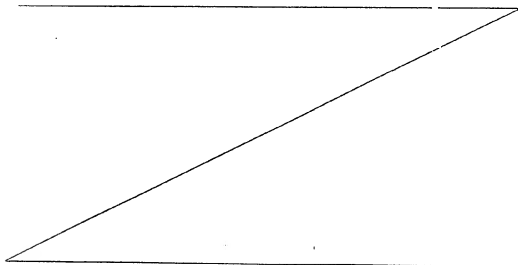
can be transformation by a plasmid or cosmid vector, or infection by a phage or viral vector etc. as known in recombinant DNA technology.

- For use as diagnostic probes the DNA of the invention which includes coding strand and/or its complement can be labelled in any conventional way, e.g. by radiolabelling, preferably with ^{32}P , enzyme labelling by the method of D.C. Ward et al., European Patent Specification 63879 or A.D.B. Malcolm et al., PCT Patent Specification WO84/03250 or fluorescently, see CNRS European Patent Specification 117,177.

The following Examples illustrate the invention. All temperatures are in $^{\circ}\text{C}$.

EXAMPLE 1

1. Selection and synthesis of an oligonucleotide primer
- A cDNA extending from approximately nucleotides 1000 to 3300 of the IBV Beaudette strain genomic RNA has been cloned in E. coli HB 101 and designated clone C5.136, see T.D.K. Brown and M.E.G. Boursnell, Virus Research 1, 15-24 (1984) and M.E.G. Boursnell, T.D.K. Brown and M.M. Binns, ib. id. 1, 303-313 (1984). The genomic map of the accompanying drawing shows C5.136 and the approximate position of a 13-base 'primer' sequence near its 5' end. This 13-base sequence is that selected for priming the synthesis of the cDNA from the spike protein coding region of IBV genomic RNA.



- 25 -

The 13-base primer sequence is located at nucleotides 256 to 268 read in the viral transcript 5'---> 3' direction. In the following partial sequence of the transcript, designated sequence formula (3), these nucleotides are underlined:

10	20	30	40	50	60
AAGAACGGTT	GGAATAATAA	AAATCCAGCA	AATTTCAAG	ATGCCCAACG	AGACAAATTG
70	80	90	100	110	120
TACTCTTGAC	TTTGAACAGT	CAGTTCAGCT	TTTAAAGAG	TATAATTTAT	TTATAACTGC
130	140	150	160	170	180
ATTCTCTGTG	TTCTTAACCA	TAATACITCA	GTATGGCTAT	GCAACAAGAA	GTAAGGTTAT
190	200	210	220	230	240
TTATACACTG	AAAATGATAG	TGTTATGGTG	CTTTGGCCCC	CTTAACATTG	CAGTAGGTTG
250	260	270	280	290	300
AATTTCAAGT	ACAT <u>ACCCAC</u>	<u>CAAACACAGG</u>	AGGTCTTGTC	GCAGCGATAA	TACTTACAGT
310	320	330	340	350	360
GTTTGGGTGT	CTGTCTTTTG	TAGGTTATTG	TATCCAGAGT	ATTAGACTCT	TTAAGCGGTG

(3)

05 The sequence of the primer was chosen on the basis of its position in the C5.136 sequence (close to the 5' terminus of the clone) and its lack of self-complementarity. Although an oligo-nucleotide sequence of only 13 nucleotides would not necessarily be unique, extensive sequencing of the entire length of C5.136
10 carried out in connection with the present invention has shown that it is unique within C5.136.

- 26 -

The primer used was the reverse complement of the above-shown 13-base sequence, i.e. was of formula (4)

5' TGTGTTTGGTGGG 3'

(4)

It was synthesised using the phosphotriester method as described by M.J. Gait et al., Nucleic Acids Research 10, 6243-6254 (1982).

05 2. Primed synthesis of ds-cDNA from viral RNA

Genomic RNA of the Beaudette strain of IBV was isolated from purified virions as described by T.D.K. Brown and M.E.G. Boursnell supra, at page 16. cDNA was synthesised from the genomic RNA using the method of U. Gubler and B.J. Hoffman, Gene 25, 263-269 (1983).

- 10 cDNA was synthesised as follows: the first strand reaction was carried out in 50 microlitres of deionised water containing 0.05M Tris-HCl pH 8.7 at 25°, 0.01M MgCl₂, 0.01M dithiothreitol, 0.004M sodium pyrophosphate, 0.001M each of dATP, dCTP, dGTP and dTTP, 40 units of human placental RNase inhibitor, 0.8 microgram of the
- 15 synthetic oligonucleotide primer described above, 10 microcuries of (alpha-³²P)-labelled dCTP, approximately 20 micrograms of the IBV genomic RNA and 160 units of AMV reverse transcriptase. The reaction mixture was incubated at 43° for 1 hour. The first strand cDNA was extracted twice with phenol/chloroform methyl
- 20 butanols (50:49:1 v/v/v), including 1 g/litre 8-hydroxyquinoline, equilibrated with 10 mM Tris-HCl pH 7.5, 1 mM EDTA and subjected to two ethanol precipitations in the presence of ammonium acetate.

- The second strand synthesis reaction mixture contained
- 25 in 100 microlitres of deionised water 0.02M Tris-HCl pH 7.5, 0.005M MgCl₂, 0.01M (NH₄)₂SO₄, 0.1M KCl, 0.15 mM beta-NAD, 0.04 mM dATP, dCTP, dGTP and dTTP, 50 micrograms bovine serum albumin, 10 microcuries of (alpha-³²P)-labelled dCTP, 22.5 units of E. coli DNA polymerase I, 10 units of RNase H, and 1 unit of E. coli DNA ligase (NAD-dependent). The reaction mixture was incubated

- 27 -

for 1 hour at 12° and then for 1 hour at 22°. The reaction mixture was then phenol/chloroform extracted and ethanol-precipitated as described above.

3. Cloning of the cDNA

- 05 dC homopolymer tails were added to the cDNA as follows: the cDNA was dissolved in 10 microlitres of 10 mM Tris-HCl pH 7.5, 1 mM EDTA and added to a final reaction volume of 50 microlitres containing 1x terminal transferase buffer (obtained from Bethesda Research Laboratories), 0.6 mM dCTP, 100 microcuries of (³H)-dCTP, 100
- 10 micrograms bovine serum albumin and 50 units of terminal transferase. The reaction mixture was incubated at 37° for one hour, heated to 65° for 10 minutes and passed over a Sepharose CL-4B column. Fractions from the leading edge of the excluded peak were pooled and ethanol-precipitated. Approximately 1 microgram of
- 15 double-stranded cDNA was obtained using this protocol. 250 ng of this cDNA was mixed with 2.5 micrograms of dG-tailed pBR322 plasmid (obtained from Bethesda Research Laboratories) and the mixture was ethanol-precipitated. The precipitate was dissolved in 40 microlitres of 0.2M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA and subjected
- 20 to the following annealing regime. It was first heated to 65° for 5 minutes, rapidly cooled to 50° and then left to cool gradually to 42° in a waterbath. The annealing was then allowed to proceed overnight to 20°.

- The annealed DNA was then transformed into E. coli strain
- 25 LE392 (see, for example, Molecular Cloning - a Laboratory Manual, T. Maniatis, E.F. Fritsch and J. Sambrook, Cold Spring Harbor Laboratory, New York, 1982) using the method of D. Hanahan, Journal of Molecular Biology 166, 557-580 (1983).

4. Isolation of a plasmid from the cloned cDNA

- 30 The E. coli LE392 transformed as described above were grown and subjected to selection for tetracycline resistance. The tetracycline-resistant colonies were screened for IBV sequences by colony hybridisation to IBV genomic RNA. Thus, the cDNA was

- 28 -

denatured and incubated with ^{32}P end-labelled alkali-treated IBV genomic RNA as the hybridisation probe. The plasmid giving the strongest signal in the colony hybridisation was designated pMB179.

5. E. coli LE392 containing plasmid pMB179 has been deposited
05 as a patent deposit under the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure on 13th June 1985 at the National Collection of Industrial Bacteria, Torry Research Station, P.O. Box 31, 135 Abbey Road, Aberdeen, Scotland AB9 8DG under the
10 number NCIB 12102.

5. Sequencing of the RNA-positive cloned cDNA

- Random subclones of pMB179 were generated by cloning either DNaseI-treated or sonicated fragments into SmaI-cut, phosphatased M13mp10 (Amersham International). Clones containing viral inserts
15 were identified by colony hybridisation with ^{32}P end-labelled alkali-treated IBV RNA or ^{32}P end-labelled reverse-transcribed viral probes. In addition PstI and RsaI fragments were cloned into PstI-digested M13mp11 and SmaI-cut, phosphatased M13mp10 respectively.
20 M13 dideoxy sequencing was carried out using (alpha- ^{35}S)dATP (Amersham International), the complete sequence being obtained on both strands. Reverse sequencing was used to obtain the last sequences required. The products of the sequencing reactions were analysed on buffer gradient gels, see M.D. Biggin et al., Proc. Natl.
25 Acad. Sci. USA 80, 3963-3965 (1983). A sonic digitiser (Graf/Bar, Science Accessories Corporation) was used to read data into a BBC microcomputer, and data was analysed on a VAX 11/750, using the programs of R. Staden, Nucleic Acids Research 10, 4731-4751 (1982) and ib. id. 12, 521-538 (1984).

30 6. Isolation of IBV S1 and S2 polypeptides

IBV, strain Beaudette, was obtained from Dr Bela Lomniczi, Budapest. All subsequent virus growth was in monolayers of primary chick kidney (CK) cells, prepared by the method of Youngner (1954).

- 29 -

The virus was passed six times in CK cells and was then plaque-purified three times. The virus in one of these plaques was passed once to produce a working stock of virus.

- For radiolabelling two 9 cm plastic dishes of CK cells were
- 05 washed twice with Eagle's minimal essential medium (EMEM) and inoculated with 4 ml of working stock plus 4 ml of fresh EMEM containing 0.2% bovine serum albumin (BSA). After 90 minutes at 37° in a 5%/95% CO₂/air atmosphere the inoculum was removed and replaced by 8 ml of EMEM. After 4.5 hours at 37° the medium was
- 10 removed and replaced with 8 ml of EMEM containing 500 microcuries of ³H-serine and 0.2% BSA. After a further 18 hours at 37° the medium was recovered, clarified, calf serum (to provide a source of protein) was added (2%) and an equal volume of saturated ammonium sulphate added. After the mixture, surrounded by melting
- 15 ice, had been stirred for 3 hours the precipitate was recovered by low speed centrifugation, dissolved in 1 ml of NET buffer (100 mM sodium chloride, 1 mM of NaEDTA, 10 mM Tris-HCl, pH 7.4) and placed on a 25-55% (w/w) sucrose gradient in NET containing 100 micrograms/ml of BSA. After centrifugation at 30,000 g average
- 20 for 16 hours at 4° the gradient was fractionated. Fractions containing virus were pooled, diluted 2.5-fold and the virus pelleted by centrifugation at 90,000 g maximum for 3 hours at 4°. The pellets were dissolved in 62.5 mM Tris-HCl pH 7.0 containing 2% SDS and 2% 2-mercaptoethanol at 100° for 2 minutes. The viral
- 25 polypeptides were separated by SDS-polyacrylamide gel electrophoresis, in a gel containing a 5-10% acrylamide gradient, using the buffers of U.K. Laemmli, Nature 227, 680-685 (1970). After electrophoresis the gel was soaked in 30 volumes of 1M sodium salicylate in water for 30 minutes. The gel was dried under
- 30 vacuum and the polypeptides located by exposure of X-ray film to the gel. The S1 polypeptide is that of highest molecular weight. The developed, dried X-ray film was placed over the dried gel and the region of the gel containing the S1 and S2 polypeptides was cut out. The polypeptides were eluted from the gel by

- 30 -

the procedure described by W.J. Welch et al., Journal of Virology 38, 968-972 (1981), extensively dialysed against distilled water containing 0.03% SDS and lyophilised. The powdered protein was dissolved in 200 microlitres of 0.1M sodium bicarbonate containing 4% SDS and added to 100 mg of p-phenylene-diisothiocyanate-treated glass (17 nm pore size) prepared by the method of E. Wachter et al., FEBS Letters 35, 97-102 (1973). Following incubation for 90 minutes at 56° under nitrogen the glass was washed with water and methanol to remove non-covalently bound material.

7. Amino acid sequencing

The glass-coupled polypeptide was then partially sequenced at the amino end by automated solid-phase Edman degradation, M. Brett and J.B.C. Findlay, Biochemical Journal 211, 661-670 (1983). The results indicated the presence in the S1 polypeptide of serine residues at positions 5, 6, 7, 14, and 20 (counting from the N-terminal end). These results unambiguously confirmed the sequence of the S1 DNA within the open reading frame. The amino acid data indicated that an 18 amino acid signal sequence MLVTPLLLVTLLCALCSA having a typical hydrophobic core and small neutral residues, alanine (A) and cysteine (C), at positions -1 and -3 from the cleavage site is cleaved from S1 during post-translational processing. The signal sequence is shown boxed in the IBV spike protein cDNA sequence formula (1) above and the region coding for the S1 N-terminus (VLYDSSSYV...) begins at nucleotide 155 in the sequence shown.

Amino acid sequencing of the S2 polypeptide indicated a serine residue at amino acid position 13 from the N-terminal end.

Two other interesting structural features of the spike precursor protein were revealed by analysis of the amino acid sequence predicted from the nucleotide sequence. Firstly, the sequence contains twenty-eight potential sites for N-glycosylation (assuming that Asn-Pro-Thr and Asn-Pro-Ser are not used) which are shown by filled circles in the sequence formula (1) above. Secondly,

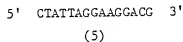
- 31 -

a hydrophilicity plot of the amino acid sequence, in the manner of J. Kyte et al., Journal of Molecular Biology 157, 105-132 (1982), showed a hydrophobic region which contains 44 non-polar amino acids preceding the charged amino acids at the carboxy-terminus of the S2 polypeptide. This hydrophobic structure probably anchors the spike protein to the viral envelope as has been proposed for similar structures on human influenza virus and fowl plague virus haemagglutinins. This region is coded for by nucleotides 3374 to 3505 and is indicated by dotted underlining in the sequence formula (1) above.

The underlined sequences at nucleotides 39 to 50 and 3556 to 3567 showing high mutual homology are the regions corresponding to the 5' ends of the bodies of mRNA E and D respectively.

EXAMPLE 2

In a procedure analogous to that of Example 1 a cDNA coding for the spike protein precursor of IBV strain M41 was prepared. The method of Example 1 was repeated using in stage (1) a 15-base primer oligonucleotide of sequence complementary to part of the sequence of the IBV Beaudette cDNA of plasmid pMB179. The primer was the reverse complement of the 15 bases numbered 3605 to 3619 in formula (1) above, i.e. was of formula (5):



Stage (4) gave rise to a plasmid pMB233 in E. coli LE392, which has also been deposited as a patent deposit under the Budapest Treaty on 13th June 1985 at the National Collection of Industrial Bacteria, under the number NCIB 12101. The IBV cDNA in this plasmid was found to extend 5'-wards from the primer for approximately 2200 base pairs.

In stage (5) sub-clones were generated from PstI-cut fragments of pMB233 in PstI-digested M13mp10, and M13 dideoxy

- 32 -

sequencing was carried out on sub-clones coding for the S1/S2 protein junction. Formula (5) below shows a partial sequence, in the region of the S1/S2 protein junction:

```

      •           ↓
      N   G   T   R   R   F   R   R   S   I   T   E   N   V   A   N   C   P
      .AATGGAACACGTCGTTTTAGACGTTCTATTACTGAAAATGTTGCAAATTGCCCC
      1690           1700           1710           1720           1730           1740

      Y   V   S   Y   G   K   F   C   I   K
      TTATGTTAGTTATGGTAAGTTTTGTATAAAA
      1750           1760           1770

```

(6)

05 which is identical with the IBV Beaudette strain cDNA sequence of formula (1). The same nomenclature is used in formula (6), the arrow denoting the 5'-end of the S2-coding region.

The entire M41 spike sequence has been inserted in two plasmids pMB 276 and pMB 250, and cloned in E. coli by a similar method to that described for pMB 233 above. Sub-clones were then
 10 made in M13mp10 as described for pMB 233. Using these clones and another clone, pMB 170, similarly prepared, the entire spike sequence of M41 was obtained. The positions of pMB 276, pMB 250 and pMB 170 relative to the Beaudette plasmid pMB 179 are shown in Figure 2 of the drawings. Plasmid pMB 250 contained a small
 15 insertion sequence of other foreign DNA shown as "IS" in Figure 2. This can readily be removed when it is desired to make a full length copy of the M41 spike sequence.

In stage (6) the S1 and S2 polypeptides of the M41 strain of IBV were isolated similarly. The virus was grown in de-embryonated
 20 chicken eggs as described by D. Cavanagh, Journal of General Virology 53, 93-101 (1981) and radiolabelled with 1 milliCurie

- 33 -

of ^3H leucine, ^3H isoleucine or ^3H valine plus 100 microcuries of ^{35}S methionine. After electrophoresis of the viral proteins in polyacrylamide gels, the gels were immediately dried under vacuum and the polypeptides located by exposure of X-ray film to the gel.

25 Partial amino acid sequence analysis of the amino-terminal of radiolabelled S2 from IBV M41 confirmed this sequence, by showing that there are isoleucine residues at positions 2 and 19 from the N-terminal, valine residues at 6 and 12, and no leucine residue in the first 20 amino acids.

10 Partial amino acid sequence analysis of S1 from IBV M41 showed a leucine residue at position 2 from the N-terminal end and a valine residue at position 9. These results are in agreement with the IBV Beaudette cDNA sequence.

Although the spike protein precursor coding cDNA of M41 appears to be highly homologous with that of Beaudette strain, there is a distinction between the two at the 3'-end. In M41 one of the nucleotides of the homology region corresponding to Beaudette 3556 to 3567 has changed. Number 3560 is a thymine base (T) instead of a guanine base (G), indicating that a stop codon UAA is present in the M41 RNA. It follows that the 3'-end of the Beaudette cDNA ends with the nucleotide sequence ..GTGGTAACT and the last 9 amino acids, at the carboxyl-terminus end of the Beaudette spike protein presursor, are not coded for in M41 strain cDNA.

25

EXAMPLE 3

This Example describes the cloning and sequencing of IBV spike cDNA of strain 6/82.

Oligodeoxynucleotides were prepared from calf thymus DNA (Sigma) by treatment with pancreatic DNase and size fractionation on DEAE-cellulose. IBV genomic RNA for strain 6/82 was prepared as described for Beaudette strain in Example 1. cDNA synthesis was carried out using the method of U. Gubler and B.J. Hoffman, supra. Thus, approximately 20 micrograms of virion RNA and 100 micrograms of calf thymus oligonucleotide primers in

- 34 -

a reaction volume of 50 microlitres (50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 10 mM DTT, 4 mM sodium pyrophosphate, 1.25 mM dNTPs were incubated with 160 units of AMV reverse transcriptase at 43°C for 30 minutes. After stopping the reaction with 20 mM EDTA

05 followed by phenol extraction, the products were precipitated with ethanol and ammonium acetate. For second-strand synthesis the products were resuspended in 100 microlitres of 20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 100 mM KCl, 0.15 mM beta-NAD, 50 micrograms ml BSA, and 40 micromolar dNTPs. 22.5 units of

10 DNA Polymerase I (Biolabs), 2.5 units of RNaseH (BRL), and 5 units of E. coli DNA ligase (Biolabs) were added to the reaction which was incubated at 12°C for 60 minutes and then at 22°C for 60 minutes. The products were phenol-extracted twice and precipitated with ethanol and ammonium acetate. Double-stranded cDNA was tailed

15 with dC residues, size-fractionated on CL Sepharose 4B, and cloned into dG-tailed PstI-cleaved pBR322. This vector was used to transform E. coli LE392 by the method of D. Hanahan, supra and selection made for tetracycline-resistant colonies. Between 2 and 4 X 10⁴ tetracycline-resistant clones were obtained in each

20 experiment of which approximately 5% were derived from uncut vector molecules. Clones were screened for the presence of viral inserts by colony hybridisation using ³²P-labelled, alkali-treated IBV 6/82 genomic RNA as a probe.

The viral inserts present in a number of clones which were

25 strongly positive in the colony hybridisation assay were tested for whether they contained IBV spike sequence, by probing with ³²P-labelled M13 sub-clones of pMB 179. Clones, pMB 252, 253 and 277, were isolated, which together encode all the 6/82 spike protein precursor (see Figure 2 of the drawings). Sub-clones in

30 M13mp10 using PstI and RsaI were made and sequenced to give the data shown in Figure 2. There are far more nucleotide changes in 6/82 than in M41, when compared to the Beaudette sequence.

- 35 -

E. coli LE392 containing plasmid pMB 252 has been deposited as a patent deposit under the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedure on 11th March 1986 at the National Collection of Industrial Bacteria, Torrey Research Station, P.O. Box 31, 135 Abbey Road, Aberdeen, Scotland AB9 8DG under the number NCIB 12221.

EXAMPLE 4

This Example illustrates the use of vaccinia virus as a vector for expression of IBV spike protein polypeptide in a mammalian cell line.

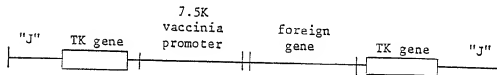
1. Insertion of IBV spike sequence into a vaccinia-compatible plasmid vector

Plasmid pMB 179 containing the IBV Beaudette spike DNA was digested with the restriction enzymes XbaI and TthIII. The restricted fragments were end-repaired with T_4 DNA polymerase using a BRL end-repair kit, and separated on a 1% agarose gel. The fragment containing the spike sequence flanked by non-coding sequences (total size 3,672 bases) was purified from agarose by the method of Dratzen et al., Analytical Biochemistry 112, 295-298 (1981). This fragment was ligated into the unique SmaI site of pGS20, a plasmid vector designed for the insertion of foreign sequences into the vaccinia virus thymidine kinase (TK) gene, described by Mackett, Smith & Moss, J. Virology, 49, 857-864 (1984). pGS20 has been widely distributed.

The following is a brief explanation of plasmid pGS20. pGS20 was constructed to contain the TK gene of vaccinia virus interrupted by (1) a vaccinia virus promoter sequence, followed immediately by (2) a sequence containing several different unique restriction endonuclease sites, whereby a foreign gene can be inserted into one of these sites.

- 36 -

The vaccinia virus promoter provides a signal for transcription of the foreign gene. When the foreign gene is inserted in pGS20, the following is the order of the various DNA sequences (shown in linear form for brevity and not to scale):



- 05 The HindIII J fragment of vaccinia virus, containing the TK gene, is interrupted by the promoter of a vaccinia virus early gene encoding a 7.5 kb polypeptide and by the foreign gene which, in the present instance, is inserted into an SmaI restriction site in pGS20. When cells are transfected with the pGS20 plasmid
- 10 containing the foreign gene and with vaccinia virus, "homologous recombination" occurs between the sites (call them A, B) on either side of the TK gene of pGS20, whereby the sequence A to B of the plasmid replaces the sequence A to B of the viral genome. Since pGS20 carries the foreign gene and a promoter, the virus
- 15 will proceed to copy the foreign gene, in this case IBV spike protein precursor cDNA. The foreign gene is then translated under the influence of its own translation initiation site. The recombinant virus-infected cells are selected for by their inability to express TK, the TK gene having been inactivated by the
- 20 insertions in it.

Following transformation of pGS20 containing the IBV Beaudette spike DNA into E. coli strain LE392, recombinant plasmids were identified by colony hybridisation to ³²P-labelled nick-translated, gel-purified IBV spike DNA fragment. DNA from

25 six of these was cut with HindIII, which cuts the spike sequence asymmetrically. One recombinant, pSB1, was selected which has the spike sequence in the correct orientation for insertion into vaccinia virus. The precise nucleotide sequence surrounding the junction between the vaccinia promoter in pGS20 and the inserted

30 IBV spike DNA fragment was determined by Maxam & Gilbert sequencing to ensure that no incorrect translational start sequences had been accidentally introduced.

- 37 -

2. Recombination into vaccinia virus

Transfection procedures and selection of recombinants were carried out as described by Mackett, Smith & Moss in "DNA Cloning: a practical approach" vol. II, ed. Glover, IRL Press Ltd., Oxford 1985, pp 191-212. Monolayers of near confluent African green monkey kidney cells, CV-1 from Flow Laboratories Inc. in 25 cm² bottles were infected with one plaque-forming unit (pfu) per cell of vaccinia virus strain WR. One hour later the cells were washed with phosphate buffered saline and then transfected with 500 microlitres per bottle of calcium phosphate-precipitated pSB1. The precipitate consisted of 20 micrograms pSB1, 1 microgram vaccinia virus DNA, 1 ml HEPES buffered saline, pH 7.12, and 50 microlitres of 2M CaCl₂ and was left on the cells for 30 minutes. Cells were harvested 2 days later and progeny viruses plaque-purified in the presence of bromodeoxy uridine (BUDR) on TK⁻143 cells available from the Wistar Institute Inc. (Other TK⁻ cells susceptible to vaccinia could be substituted). The TK⁻ selected viruses were grown up in small monolayers of TK⁻ cells and screened for the presence of spike sequences by dot-blotting onto nitrocellulose and probing with ³²P-labelled nick-translated pMB179. Two positive recombinants, vaccinia-SP1 and vaccinia-SP2 were plaque-purified again on TK⁻ monolayers with BUDR selection, re-screened by dot-blotting then large stocks of vaccinia-SP1 were grown up in CV-1 cells without selective conditions. Vaccinia-SP1 was purified by twice banding in 36-50% w/v sucrose gradients and DNA was extracted from virions. This DNA was cut with HindIII and the resulting fragments run out on a 0.6% agarose gel. Ethidium bromide staining and UV visualisation of the DNA indicated that the 5 kb HindIII J fragment of wild-type DNA (containing the vaccinia TK gene) was absent from the recombinant vaccinia-SP1 and instead there were two new HindIII fragments, the sizes of which were consistent with the insertion into vaccinia TK of the IBV

- 38 -

Beaudette spike sequence. Southern blotting of this agarose gel and probing with nick-translated 32 P-labelled pMB179 confirmed that these new fragments did indeed contain the spike sequence.

3. Expression of IBV spike protein polypeptide in

05 monkey kidney cells

CV-1 cells in 25 cm² bottles were infected with 40 pfu per cell of wild type or vaccinia-SP1 virus and radiolabelled between 2 and 6 hours post infection with 80 microcuries of 35 S-methionine. Lysates were prepared from infected and control cells at 6 hours after infection and immunoprecipitated with rabbit anti-spike protein serum and staphylococcal protein A as described by Mackett, Smith & Moss 1985, loc. cit. The precipitated polypeptides were separated by polyacrylamide gel electrophoresis and visualised by autoradiography. In lysates prepared from vaccinia-SP1 infected cells, two high molecular weight polypeptides were specifically precipitated by anti-spike protein serum which were consistent in size with spike proteins S1 and S2 of IBV. These were absent from the cell lysates of uninfected and vaccinia wild type-infected cells. Indirect immunofluorescent antibody staining of surface fixed vaccinia-SP1 infected cells was carried out using rabbit anti-spike protein serum and fluorescein conjugated anti-rabbit serum as described by Mackett, Smith & Moss. Strong surface labelling consistent with the spike polypeptide being expressed at the cell membrane of vaccinia-SP1 infected monkey kidney cells was observed.

- 39 -

CLAIMS

1. A DNA molecule which codes for an IBV spike protein polypeptide comprising a S1 or S2 polypeptide or for an antigenically determinant polypeptide thereof.
2. A DNA molecule according to Claim 1 wherein said polypeptide
15 has at least 80% amino acid sequence homology with the corresponding polypeptide of IBV Beaudette, M41 or 6/82 strain.
3. A DNA molecule according to Claim 2 wherein the polypeptide has at least 90% said amino acid sequence homology.
4. A DNA molecule according to Claim 1, 2 or 3, comprising a
10 nucleotide sequence which codes substantially only for any of (1) the spike protein precursor, (2) the S1 signal plus the S1 polypeptide, (3) the S1 polypeptide and (4) the S1 polypeptide plus the S2 polypeptide, each of which sequences can be truncated by a
15 sequence of up to 30 nucleotides at either or both ends and/or flanked by up to 100 nucleotides of contiguous IBV cDNA or by any length of a foreign DNA sequence, at either or both ends.
5. A DNA molecule according to Claim 4 wherein any said flanking is by up to 20 nucleotides of contiguous IBV spike protein cDNA and any said truncation is by a sequence of up to 15 nucleotides.
- 20 6. A DNA molecule according to any preceding claim wherein the polypeptide has at least 75% nucleotide sequence homology with the corresponding sequence of IBV Beaudette, M41 or 6/82 strain.
7. A DNA molecule according to Claim 6 wherein the nucleotide sequence homology is at least 80%.
- 25 8. A vector carrying an inserted sequence of a DNA molecule claimed in any one of Claims 1 to 7.
9. A vector according to Claim 8 which is a poxvirus vector comprising a viral promotor sequence linked to an inserted
sequence of a DNA molecule claimed in any one of Claims 1 to 7.
- 30 10. A vector according to Claim 9 wherein the virus is fowlpox virus.
11. A vector according to Claim 8 which is a cloning vector.

- 40 -

12. Mammalian cells containing a DNA molecule claimed in any one of Claims 1 to 7.
13. Mammalian cells according to Claim 12 wherein the DNA molecule is contained in a vector defined in Claim 9 or 10.
- 05 14. A host incorporating a cloning vector defined in Claim 11.
15. A host according to Claim 14 incorporating a plasmid containing the IBV spike protein precursor coding cDNA, said cDNA being present in patent deposit NCIB 12101, 12102 or 12221, or showing at least 90% nucleotide homology therewith.
- 10 16. A host according to Claim 15 wherein the degree of homology shown is at least 95% nucleotide homology.
17. A host according to Claim 15 or 16 which is an E. coli bacterium.
18. Artificial IBV spike protein polypeptide comprising a S1 or S2 polypeptide or an antigenically determinant polypeptide thereof.
- 15

1/2

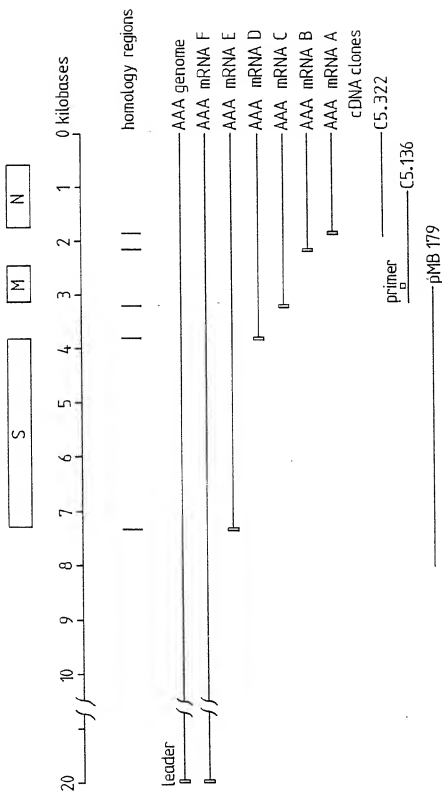


Fig. 1

2/2

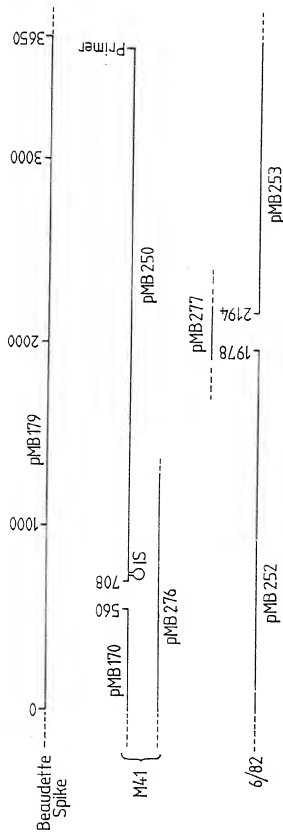


Fig. 2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/GB 86/00181

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC⁴: C 12 N 15/00; A 61 K 39/215; C 12 N 5/00; C 07 K 13/00

II. FIELDS SEARCHED

Minimum Documentation Searched *

Classification System

Classification Symbols

IPC⁴

C 12 N
A 61 K

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched *

III. DOCUMENTS CONSIDERED TO BE RELEVANT *

Category *	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	Modern Approaches to Vaccines, Cold Spring Harbor Laboratory, 1984, New York, (US) D. Cavanagh et al.: "Genetically engineered vaccine against avian infectious bronchitis virus with the advantages of current live and killed vaccines", pages 215-218, see page 217, paragraph 2 (cited in the application)	1-8
A	Virus Research, volume 1, 1984, Elsevier T.D.K. Brown et al.: "Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries", pages 15-24, see figure 1 (cited in the application)	1-8
A	Gene, volume 29, July/August 1984, Elsevier M.E.G. Boursnell et al.: "Sequence of coronavirus IBV genomic RNA: a 195-base open reading frame encoded by mRNA B, see the whole article (cited in the application)	1-8

* Special categories of cited documents: 10-

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (see specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"A" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search
2nd July 1986

Date of Mailing of this International Search Report

31 JUL 1986

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

M. VAN MOL

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	Virus Research, volume 1, 1984, Elsevier M.E.G. Boursnell et al.: "Sequence of the membrane protein gene from avian corona- virus I.B.V.", pages 303-313, see the whole article (cited in the application) ---	1-8
X, P	Journal of General Virology, volume 66, no. 4, April 1985, (GB) M.M. Binns et al.: "Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV", pages 719-726	1-8
